Using Standardization and Ontology to Enhance Data Protection and Intelligent Analysis of Electronic Evidence

Eoghan Casey (University of Lausanne)[1]*, Maria Angela Biasiotti and Fabrizio Turchi (Institute of Legal Information Theory and Techniques of the Italian National Research Council)

1) Introduction

There are increasing amounts of personal information being stored within various systems and organizations to support health care, financial transactions, telecommunications, and many other necessities of modern life. A limited portion of this information may be required as part of an authorized digital inquiry, including criminal, civil, and regulatory matters. At every phase of a digital inquiry, information must be well organized and the provenance of electronic evidence must be maintained for forensic purposes. Furthermore, electronic evidence must be carefully protected to prevent privacy violations, exposure of secrets, and violation of license agreements.

Fulfilling investigative and forensic requirements, while organizing data and protecting information appropriately, is supported by the open community-developed specification language called CASE, the Cyber-investigation Analysis Standard Expression (see [6]). This standardisation effort is a rational progression from the foundational work on Digital Forensic Analysis eXpression (DFAX) which was created by many of the same contributors as CASE (see [5]). CASE combines the lessons learned from developing DFAX with the experiences of related efforts, including Hansken[2] (see [3]). To further promote a common structure, CASE aligns with and extends the Unified Cyber Ontology (UCO), which provides a construct for representing cyber artifacts across multiple domains, including criminal justice, digital forensic science, incident response, eDiscovery, and regulatory compliance.[3]

One of the primary aims of CASE is to provide investigative entities and other involved parties with a standard way to represent and exchange electronic evidence. CASE includes the capability to mark specific information using any level of sensitivity, including sensitive, private, proprietary, and public. By enabling efficient and controlled sharing of electronic evidence between different parties and jurisdictions, CASE supports the following use cases:

- Enhance coordination between criminal justice agencies both nationally and internationally within legal frameworks such as the European Investigation Order (EIO) and mutual legal assistance (MLA) procedures.
- Restrict access to privileged, proprietary, and personal in criminal and civil matters.
- Control disclosure of information in regulatory compliance inquiries.
- Enable vendors to restrict use of data that is covered under license agreements.
- Enable users to mark their own data as private in systems maintained by government organizations and service providers.

In addition, using CASE to represent electronically stored information provides a firm foundation for advanced forms of data analysis and exploration. This standard provides structure for capturing details about the context of cyber-information, including links between items and how data were handled, transferred, processed, analyzed, and interpreted. The increased organization of data, metadata, and their context that is encouraged through CASE provides an enriched latticework of information, opening new opportunities for contextual analysis, pattern recognition, machine learning, visualization, and data protection mechanisms. Data markings are an integral component of CASE, permitting information to be labeled as private or sensitive, and to be shared or protected appropriately at different levels of trust and classification.

This paper explains how the CASE standard supports automated information exchange, including the use of data markings to specify how information can be treated and shared, and how CASE creates opportunities for more sophisticated analysis techniques.

2) Challenges in handling electronic evidence

There are many different types of electronic evidence, and new kinds are created by technological advances. To clarify the full scope of electronic evidence, the EVIDENCE Project[4] defined three categories:

1. Physical or traditional (non-electronic) evidence such as a murder weapon, fingerprint, or bloodstain that has been digitized (e.g., digital photographs of a murder weapon).
2. Analogical evidence: evidence originally in an analogue form (e.g., paper documents, videotape, vinyl) that has been converted to digital format.
3. Digital evidence: information of probative value originally created, stored, or transmitted in binary form by any computing device (see [2]).

The CASE standard can be used to represent all three of these categories throughout the electronic evidence lifecycle.

* Corresponding author. Tel.: +41-21-692-46-12; fax: +41-21-692-46-05, *E-mail address*: eoghan.casey@unil.ch.

[2] Hansken is a service-based automated system developed by the Netherlands Forensic Institute (NFI) for processing large amounts of digital evidence.

[3] For a comprehensive overview of the Unified Cyber Ontology and how it is used by CASE see [7].

[4] Evidence Project – "European Informatics Data Exchange Framework for Courts and Evidence," www.evidenceproject.eu.

2.1) The electronic evidence life cycle

The process of handling electronic evidence can be divided into several phases, and there may be a need to exchange information during any of these phases. The first phase includes the identification, collection and anti-contamination precautions (searching the scene, collecting the evidence, packaging and labelling and creating documents reporting the activities performed at every step) of electronic evidence. In the second phase, the acquisition of the source of evidence takes place, determining which items are most likely to serve the purposes of the investigation, which are the most time sensitive, which are most at risk of being lost or corrupted, including the identification of similar issues. During the third phase, the findings are evaluated and interpreted. The fourth phase includes the presentation of the results in a report, which should include factual findings, interpretation, and expert opinion. The report and presentation are essential steps in the electronic evidence lifecycle, because the court will examine the report that should contain all relevant findings as well as technical and non-technical explanations of the case and its issues.

When electronic evidence is acquired and exchanged, information about its provenance must be maintained to help establish authenticity and trustworthiness. This information includes the origin of the evidence, its condition, and any unique or distinctive characteristics. Continuity of possession is a key component of provenance, any irregularities in the handling of electronic evidence could spoil the probative value of the evidence, potentially making it inadmissible (see [17]). Provenance also includes information used to establish integrity, i.e., that electronic evidence has not been altered since the time it was created, stored, or transmitted. Provenance can also include "the set of tools and transformations that led from acquired raw data to the resulting product" (see [16]). CASE can be used to represent all of these details in digital form, combining provenance information with electronic evidence into a cohesive and comprehensive evidential package to support subsequent processes.

One of the primary challenges associated with electronic evidence is the lack of uniformity in legislation and criminal procedures in different countries. Some countries have clearly defined rules as to admissibility of evidence in legal proceedings, while in other countries admissibility is flexible. Privacy and data protection laws in some countries prevent the collection of certain electronic evidence, and have varying data retention periods. Legislation may furthermore not sufficiently address the realities of modern investigations, especially when it comes to evolving new technologies. The CASE standard is designed with these challenges in mind, providing sufficient flexibility within the structure of the underlying Unified Cyber Ontology (UCO) to accommodate variations in legal requirements and criminal procedures across different jurisdictions.

2.2) Electronic evidence exchange

Currently, the primary mechanism for exchanging evidence is through regular international procedures for mutual legal assistance in criminal matters, but these procedures are time-consuming, unpredictable and not well-equipped to deal with the electronic evidence. Moreover, in cross-borders criminal cases, cooperation is mostly human based.

The present situation raises the following issues:

- Exchanging evidence may be slow, which can create significant problems in time-sensitive matters;
- Exchanging evidence may be expensive, e.g., personnel travelling between countries to transport the items being exchanged;
- Judicial and police authorities must invest substantial resources to keep pace with the development of forensics technology;
- Exchanging trusted procedures are of utmost importance.

To address the challenges associated with efficient and secure exchange of electronic evidence between European countries, the Evidence project defined the following functional requirements:

- A standard for representing data and metadata that comprise electronic evidence, including actions (i.e. tasks), participants (e.g. subjects, victims, authorities, attorneys, investigators, forensic analysts), tools (i.e. hardware and software for performing forensics processes), digital and physical objects involved in the investigative case (e.g. hard disks, mobile devices, files, messages), and the relationships between objects (e.g. stored-on, contained-within, decoded-from).
- A standard for representing provenance information associated with electronic evidence, including information about who (performer or tool) did what (e.g., seized, imaged, executed), when, and how.
- A standard for representing the results of the forensics tools.
- A standard that supports any data markings agreed among sharing parties to specify obligations and controls on information that is being exchanged.
- A secure and trusted platform for exchanging electronic evidence.[5]

In addition, the following questions must be taken into consideration when establishing the exchange process:

(i) What information should be exchanged?

[5] Preferably extending and integrating with secure platforms that are already in place (e.g., Siena led by Europol, e-Testa led by Eurojust, I-24/7 led by Interpol).

(ii) When may the exchange take place?
(iii) How can the information be exchanged, taking into consideration security/privacy issues?
(iv) How to manage the exchange of extremely large amounts of electronic evidence? Which stakeholders are involved?

The CASE standard has been developed to satisfy many of the above requirements, including being platform independent, and CASE is being adopted by developers of existing systems that are used to handle electronic evidence.

3) Evolution of standards for representing and exchanging electronic evidence

The need for a standard to represent and exchange the full scope of electronic evidence has been heightened by the growing relevance of this form of evidence in a wide range of situations.

Existing standards for exchanging general criminal justice information, including the National Information Exchange Model (NIEM), have not kept pace with the evolution of electronic evidence (https://www.niem.gov).

Prior work to represent digital evidence in a common format was focused on specific forms of digital evidence and did not encompass the full scope of electronic evidence and cyber-investigation information. Digital Forensics XML (DFXML) represents file system information (see [8], [11] and [12]). The Advanced Forensic Format (AFF4) stores digital evidence using the Resource Description Framework (see [7]). XIRAF was developed by the NFI and has been superseded by their Hansken data model (see [1] and [4]). The XML Data Encoding Specification for Intelligence Document and Media Exploitation (DOMEX) was developed by the U.S. government to share certain types of information, including a limited set of mobile device details (see [18]). The use of a UMF (Universal Message Format) to exchange electronic evidence in the SIENA system provides tags for representing identities, relationships, and activities.

The value of using various standards for exchanging information related to cyber-security incidents has been emphasized by European Union Agency for Network and Information Security (see [9] and [10]).

Representation of cyber threat intelligence (CTI) information has been standardized using STIX.[6] The focus of STIX Observables which are embedded within and dependent on CTI context-specific structure of the STIX schema makes it unsuitable as a foundation for the broader scope of electronic evidence and cyber-investigation.

Overall, the existing formats described above have limited expressivity for characterizing cyber observables, only keep track of limited provenance information, and lack a supporting ontology.

3.1) Digital Forensic Analysis eXpression (DFAX)

To address the full scope of requirements for representing and exchanging electronic evidence in a standardized format, the Digital Forensic Analysis eXpression (DFAX) was initially developed by Casey, Barnum, and Back (see [5]).[7] DFAX and a supporting ontology called the Unified Cyber Ontology (UCO), defined structures to represent procedural aspects of the digital forensic domain, including chain of custody, case management, and forensic processing. DFAX leveraged CybOX (Cyber Observable eXpression) to represent cyber objects along with their interrelationships in order to support standardized storing and sharing certain cyber-related information.[8] DFAX provided an open-source standardized format to represent digital observables, interrelationships, and provenance, as well as associated actions and context. In addition to representing static information within electronic evidence such as a file with a specific MD5 hash value, DFAX could be used to represented dynamic activities such as a Registry key being created or a file being deleted.

The main item of DFAX structure is the DFAX Package (called a *Compilation* in CASE) that can group together many different complex elements, including three reference elements, as shown in Figure 1. Each complex element includes further sub elements.

[6] Structured Threat Information eXpression (http://stix.mitre.org/)
[7] Originating from Casey, Barnum and Back at MITRE. MITRE is a not-for-profit organization that operates research and development centers sponsored by the federal government (www.mitre.org).
[8] Development of CybOX as an independent standard ceased in 2016. CybOX was replaced by STIX Observable as an integrated component of STIX, and concentrates on representing cyber threat intelligence.

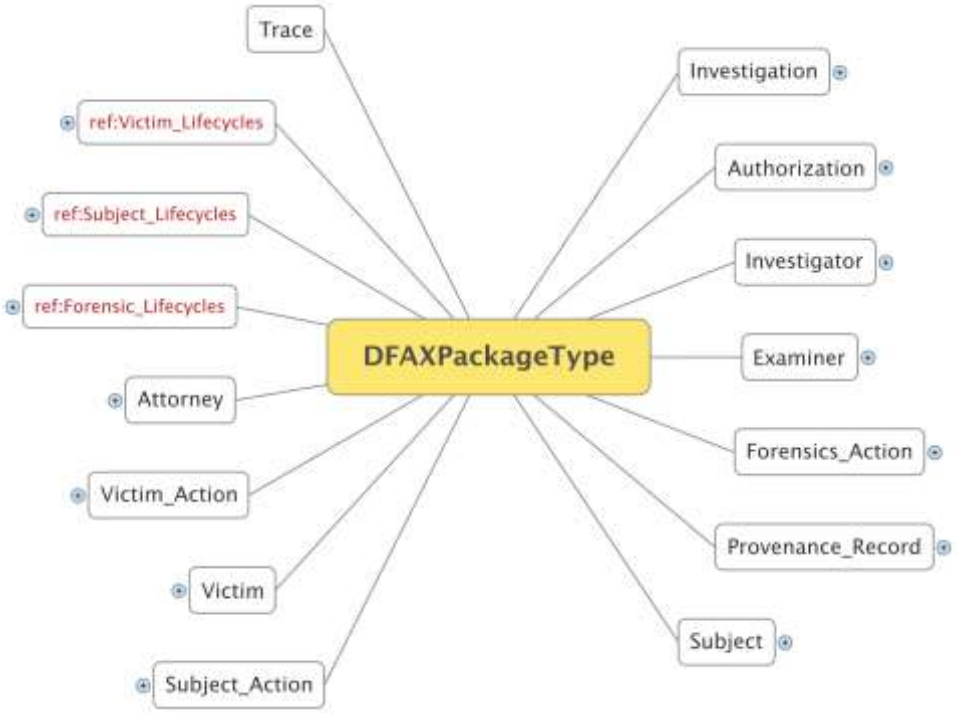

***Figure 1**: Simplified depiction of the objects that can be grouped together and related to each other within a DFAX Package (called a Compilation in CASE).*

Extended community involvement, combined with the deprecation of CybOX, caused DFAX to evolve into the CASE standard, which incorporates lessons learned from DFAX, CybOX, DFXML, XIRAF, and Hansken.

3.2) Cyber-investigation Analysis Standard Expression (CASE)

CASE has structures to represent technical objects (Traces) and their properties as well as the more procedural aspects of the cyber-investigation and digital forensic domains, including those for chain of custody, case management, forensic processing. CASE has been developed with extensibility in mind: new object types and associated properties can be added without altering the core schema. This core robustness is supported by the Unified Cyber Ontology (UCO) which provides domain (e.g. Action Lifecycle) and represents the actors/stakeholders, or Identities, involved in a case, such as: Attorney, Investigator, Examiner, Subject, Victim. These identities can be defined depending on the context, and using an existing standard that CASE references. An open question is what existing standard to use for representing Identities, such as the NIEM or OASIS Customer Information Quality (CIQ) standards. In addition, Observables have been deprecated and replaced by Traces in CASE, which relates to the more general Cyber-Item in UCO.

4) Cyber-investigation example

A practical example is provided here to demonstrate how CASE can be used to represent electronic evidence and associated cyber-investigation information. This example describes all of the objects, electronic evidence, people, places, and activities in an investigation involving a personal computer, a USB drive, and a smartphone. The investigative process is described as the sequence of actions related to identification, acquisition, and analysis phases of the electronic evidence lifecycle, and provides an output file in a standard format containing all the information related to the case. The case is described as follows:

> *The CEO (Rex Cox) of a private company (EarlyBird Corporation) is suspected of trade secrets infringement and disclosure of private information to another company (NightOwl Corporation), the forensic lifecycle is made by Identification, Acquisition, Analysis and Report*

***Example 1**: Cyber-investigation scenario to illustrate the lifecycle of electronic evidence as shown in Figure 2.*

The outline of the case is represented in Figure 2 with high level CASE objects and entities on the left (orange), CASE property level information on the right (yellow), and a list of Forensic Actions at the bottom. Each Forensic Action is represented separately using CASE, including references to the objects that were acted upon.

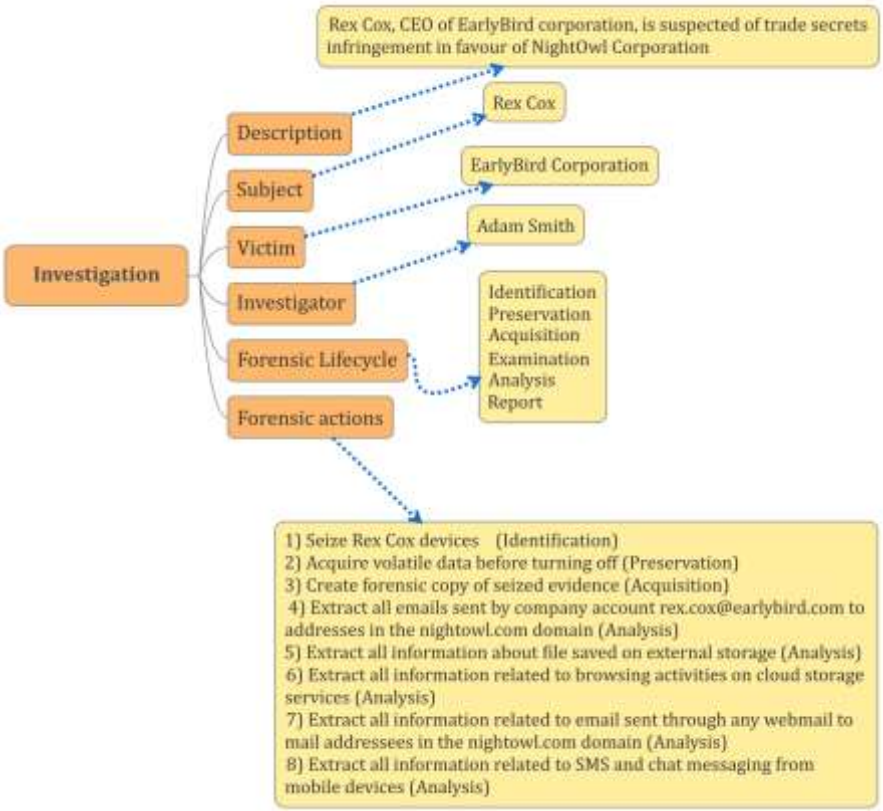

***Figure 2**: Depiction of how CASE can be used to represent all aspects of an investigation with high level CASE objects and entities on the left (orange), CASE property level information on the right (yellow), and a list of Forensic Actions at the bottom. Each Forensic Action is represented separately using CASE, including references to the objects that were acted upon.*

4.1) Resolution of issues from DFAX to CASE

Development of the community-driven CASE standard is ongoing, and there are open issues being addressed, taking into account typical scenarios that investigators and analysts encounter during real cases. Some issues that were encountered while developing DFAX are described here, along with an explanation of how they were addressed in CASE.

a) In DFAX there was a strict structure representing the roles of Authority – Investigator – Victim – Subject. These entities were discussed with experts and it was determined that greater flexibility was needed to accommodate different contexts and legal systems. Addressing this limitation, CASE and UCO provide a flexible structure that can be used to represent any identity and role.

b) DFAX used CybOX Objects which focuses on cyber-threat intelligence rather, and did not have strong support for objects in other related domains, including cyber-investigation and digital forensics. To address this issue, CASE adopted the approach used in the Hansken system to create a flexible data model (based on duck testing[9]) that can be easily extended to represent any cyber-information and its properties.

c) The original design of the standard used the concept of Provenance_Record as the Input/Output of every single action, which was effective in the majority of use cases. However, there are cases where the description of the Provenance_Record in terms of Traces is not a good fit. To address this issue, CASE is being developed to accommodate different use cases, including a single Provenance_Record for a group of objects that were extracted together, rather than each object having an individual Provenance_Record.

d) In DFAX, there is no easy way to describe deleted content that can be carved from the unallocated space or recovered through the file system of the disk. There was a shortage of objects useful for describing elements that should be fruitful from a cyber-investigation or digital forensic point of view. CASE addresses these limitations, and standard representations of additional object types are being developed specifically to address the requirements of cyber-investigations and digital forensics.

As more developers become involved in this community-driven initiative, the CASE/UCO structure is being refined to facilitate implementation in tools and systems that support cyber-investigations.

[9] Duck typing allows data to be defined by its inherent characteristics rather than enforcing strict data typing. CASE objects can be assigned any rational combination of property bundles, such as a file that is an image and a thumbnail. When employing this approach, data types are evaluated with the duck test, which uses inference to the best explanation (a.k.a. abductive reasoning). Simply stated, if it walks like a duck, swims like a duck, quacks like a duck, and looks like a duck, then it probably is a duck.

5) Formalizing obligations and controls for information sharing

A crucial aspect of information exchange is being able to specify the permitted conditions for sharing and to enforce exchange policies. UCO provides for data markings that CASE can use to support proper handling of shared information. Any marking mechanism can be employed, including Traffic Light Protocol (TLP)[10] and organization specific data marking vocabularies. For instance, the Forum of Incident Response and Security Teams (FIRST) has an Information Exchange Policy (IEP) with defined vocabulary for controlling how information is shared, which includes TLP (https://www.first.org/iep). Example 2 demonstrates how CASE is used to implement FIRST's IEP (version 1.0) specification.[11]

```
{
  "@id": "marking-01bc4353-4829-4d55-8d52-0ab7e0790df9",
  "@type": "MarkingDefinition",
  "name": "FIRST.org Mailing List IEP",
  "definitionType": "IEPMarking",
  "definition": [
    {
      "@type": "IEPMarking",
      "version": 1,
      "reference": "https://www.first.org/mailing-list-iep",
      "start--date": "2016--06--09 10:09:00",
      "end--date": "2016--12--31 10:09:00",
      "encrypt--in--transit": "MAY",
      "encrypt--at--rest": "MAY",
      "permitted--actions": "EXTERNALLY VISIBLE DIRECT ACTIONS",
      "affected--party--notifications": "MAY",
      "tlp": "AMBER",
      "attribution": "MUST NOT",
      "obfuscate--affected--parties": "MUST",
      "unmodified--resale": "MUST NOT",
      "external--reference": "https://www.first.org/about/policies/bylaws"
    }
  ]
}
```

***Example 2**: Example of FIRST's Information Exchange Policy (IEF) implemented using CASE, mapping the example in Appendix A of FIRST IEF Version 1 (https://www.first.org/iep#appendix-a).*

These data markings can then be applied to any information that is being shared. For instance, the data markings in Example 2 can be specified for particular elements that require special protections, or can be referenced in the header of a shared CASE bundle to apply to all of the information being shared. In Example 3, the data markings in Example 2 are applied to an individual Trace object, and in Example 4, these data markings are applied to a complete Bundle of content.

```
{
  "@id": "email-59e9cf76-08c3-4f0b-a319-2a3b55b54f03",
  "@type": "Trace",
  "objectMarking": ["marking-01bc435348294d558d520ab7e0790df9"],
  "propertyBundle": [
    {
      "@type": "EmailMessage",
      "to": ["EmailAccount-bb704188-de16-4743-92fc-b4cba6f9f464"],
      "cc": ["EmailAccount-6c0e2c89-05c2-4713-8a2e-51126725c783"],
      "bcc": ["EmailAccount-a41737ad-558c-44a4-8031-40c623b3f07b"],
      "from": "EmailAccount-bcc67257-331c-4151-8818-1196eb91e7e0",
      "subject": "Example email message",
      "sender": "EmailAccount-bcc67257-331c-4151-8818-1196eb91e7e0",
      "receivedTime": "2017-03-28T13:44:23.40Z",
      "sentTime": "2017-03-28T13:44:22.19Z",
      "messageID": "CAKBqNfyKo+ZXtkz6DUjW-pvHkJy6kwO82jTbkNA@mail.gmail.com"
    }
  ]
},
```

***Example 3**: Example of a single Trace object (content data encoded in base64) represented using CASE with the data markings asserted in Example 2 applied to the entire Trace object.*

[10] The Traffic Light Protocol uses four color designators to specify how information is permitted to be shared: RED (no sharing with parties outside the direct exchange), AMBER (only share within the organization of sharing parties), GREEN (only share among peers and partner organizations within a given community or sector, do not share publically), and WHITE (share freely, subject to standard copyright rules). TLP does not address further conditions or restrictions such as encryption rules, license agreements, and permitted actions (see https://www.us-cert.gov/tlp).

[11] In technical terms, the IEPMarking type would simply be defined as an extension of the MarkingModel class in UCO, which is implemented in CASE. In this way, CASE can implement any data marking specification that is agreed among the parties that are exchanging information.

```
{
  "@id": "bundle-b0d719c0-db84-4a41-9026-9caf71a5720c",
  "@type": "Bundle",
  "objectMarking": ["marking-01bc4353-4829-4d55-8d52-0ab7e0790df9"],
  "content": [
    {
      "@id": "phoneaccount-aba00444-9f1a-44f2-81ba-1be4f7a27314",
      "@type": "Trace",
      "propertyBundle": [
        {
          "@type": "Account",
          "accountIssuer": "AT&T",
          "isActive": true
        },
        {
          "@type": "PhoneAccount",
          "phoneNumber": "1237771337"
        }
      ]
    },
    {
      "@id": "phoneaccount-6efe9553-bd61-486f-ae9e-5f9dfe613a7f",
      "@type": "Trace",
      "propertyBundle": [
        {
          "@type": "Account",
          "accountIssuer": "Sprint",
          "isActive": true
        },
        {
          "@type": "PhoneAccount",
          "identifier": "1234560000"
        }
      ]
    },
    {
      "@id": "phonecall-5c76372a-5bce-4f73-8800-91691d8f0b57",
      "@type": "Trace",
      "propertyBundle": [
        {
          "@type": "PhoneCall",
          "callType": "mobile",
          "startTime": "2010-01-15T17:59:43.25Z",
          "endTime": "2010-01-15T18:30:41.25Z",
          "from": "phoneaccount-aba00444-9f1a-44f2-81ba-1be4f7a27314",
          "to": "phoneaccount-6efe9553-bd61-486f-ae9e-5f9dfe613a7f",
          "duration": "PT31M2S"
        }
      ]
    }
```


Example 4*: Example of a Bundle of CASE content with the data markings asserted in Example 2 applied to the entire bundle of information including its content objects.*

Processes for adding and enforcing data markings can be automated to facilitate appropriate handling of shared information.

5.1) Enforcing information sharing obligations and controls

To automate information sharing, including enforcement of data markings, it is necessary to define the services and messages that will carry content between sharing parties. For instance, TAXII[12] specifies a protocol for exchanging information over HTTP/HTTPS, and provides several exchange methods (push, pull, query). TAXII supports digital signatures so that information being shared can be authenticated and its integrity can be verified.

A nuance of using digital signatures in this context is the need to share subsets of the originally signed information. For instance, when some information within a bundle is restricted and must be redacted, it is necessary to generate a new digital signature on the subset of information that can be shared. Rather than require the originator to repeat the process of digitally signing each subset of sharable information, this process can be automated within the exchange system. For example, when it is necessary to extract a subset of information for sharing, the exchange system can automatically verify

[12] Threat Automated eXchange of Indicator Information (http://taxii.mitre.org/)

the contents of the subset against the original information, and then automatically sign the subset of information using an authorizer server certificate/key.

The cyber threat intelligence community uses TAXII to support automated exchange of information in STIX format. TAXII can also be used to exchange other types of information, including alerts formatted using the Common Alerting Protocol (CAP). TAXII could be adapted for use in electronic evidence exchange systems, transporting CASE bundles between sharing parties. Alternative mechanisms for controlling how the information is exchanged and accessed could be developed by vendors, organizations, or the community as a whole (e.g., using a combination of encryption and licensing). Any of these exchange mechanisms will be able to carry information represented using the CASE standard.

5.2) Keeping track of how electronically stored information is treated

An important aspect of information sharing and data protection is transparency and auditability. The CASE standard supports these requirements by providing the structure necessary to keep track of how electronically stored information is handled and processed (a.k.a. provenance). This provenance information includes the who, what, where, when, how, and sometimes why of how electronic evidence was treated.

For instance, processing a mobile device such as acquiring a forensic copy can be represented by Forensic Actions, using the Provenance Record as the input object in order to keep track of each step in the provenance (see Example 5).

```
{
  "@id": "forensic_action1",
  "@type": "ForensicAction",
  "name": "preserved",
  "startTime": "2017-04-05T07:36:24.35Z",
  "propertyBundle”: [
    {
      "@type": "ActionReferences",
      "instrument": ["warrant1"],
      "location": "crime_location1",
      "performer": "investigator2",
      "object": [
        "mobile_device1"
      ],
      "result": [
        "provenance_record13"
      ]
    }
  ]
},
{
  "@id": "annotation14",
  "@type": "Annotation",
  "tag": ["forensic"],
  "description": "Forensic preservation of mobile device.",
  "object": [
    "forensic_action13"
  ]
},
{
  "@id": "provenance_record13",
  "@type": "ProvenanceRecord",
  "description": "Samsung Galaxy Edge",
  "exhibitNumber": "CAPD-2017011601",
  "object": "mobile_device1"
}
```

Example 5*: Example of a Forensic Action and Provenance Record.*

Authorizations for cyber-investigations such as search warrants and court orders are represented in CASE as Traces, which are referenced within Forensic Actions as shown in bold in Example above. This approach allows a single authorization to be referenced from multiple Forensic Actions without duplicating the information.

The environment in which the Forensic Action occurred can be represented, whether it be a computer system (forensic tool running on a laptop running Windows 10), or a physical location (photographing evidence in an office).

The role of the Performer can be specified within each Forensic Action, allowing one person to have multiple roles throughout the forensic lifecycle: a first responder during the preservation phase, a forensic examiner during the examination phase, and an expert witness during the presentation phase.

Using provenance information, it possible to evaluate the reliability of information and provide transparency. If information is incorrect or has been mishandled or insufficiently protected, the provenance information can help determine where the problem originated. In addition, provenance information can help track down and amend subsequent activities that utilized unreliable information, thus limiting long term repetition of bad information.

6) Organizing and analyzing electronically stored information
In addition to representing individual traces, it is important to capture their context and relationship with other traces and entities, for provenance and investigative purposes. CASE represents the linkages between items using a combination of embedded references for properties that cannot change and relationships for things that can change or are externally asserted. The use of JSON-LD to represent linked data enables automated enrichment using additional sources such as identity lookups, word translation, and geolocation databases.

Being able to represent structure by defining relationships within the data enables search and analysis methods at a higher level of abstraction, including graph query and pattern matching. For instance, defined relationships between items as shown in Figure 3 could be utilized to perform a graph search for all email messages with a picture attachment from the subject to the victim (see [5]).

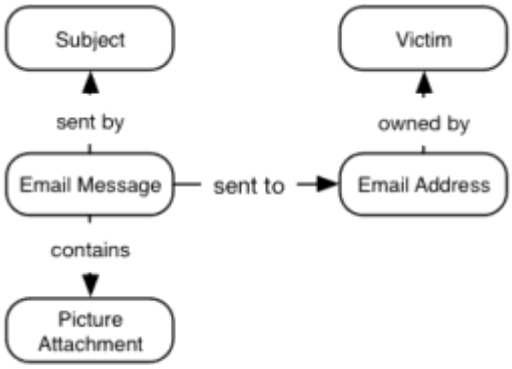

***Figure 3**: Depiction of relationships (graph).*

The UCO provides an ontology that generalizes how each of these types of information is structured (among many others), and can be useful across multiple domains, including digital forensics, incident response, and counterterrorism.

Conclusions
The primary motivation for CASE is to enable sharing of electronically stored data in a standardized and secure manner. CASE and the supporting Unified Cyber Ontology, was designed from the outset to support data protection and privacy. As CASE becomes more widely adopted by systems and tools that process and store electronic information, there is greater potential for automated normalization, combination correlation, and validation of information. Data markings in CASE are used to specify restrictions on sharing information as well as how information should be treated. To automate the application and enforcement of data markings, it is necessary to update tools and systems that support electronic evidence exchange using a standard such as TAXII. The EVIDENCE project is collaborating with organizations that have existing infrastructure connecting criminal justice organizations to enable the flow of standardized electronic evidence in a manner that properly enforces data markings to ensure that security, privacy, and proprietary information are protected.

Another primary motivation for CASE is to enable more advanced and comprehensive correlation and analysis, searching for things not strings. In addition to fusing together disparate sources of information, CASE expresses information in a fully structured form that supports a multitude of analysis methods. In addition to searching for specific keywords or characteristics within a single case or across multiple investigations, having a structured representation of cyber-investigation information allows pattern searching, graph query, data mining, and other sophisticated analytics. Improved capabilities to find important items can help solve an investigation, and more effective approaches to finding non-obvious similarities between investigations can help discover links between related or similar activities. This structure also facilitates data abstraction and visualization, providing a solid foundation for more powerful analysis support systems.

The initial draft of CASE has been released for broader community use and development (https://github.com/casework) along with the supporting UCO (https://github.com/ucoproject). Community development is ongoing to expand the types of information that CASE can represent and to provide more comprehensive and refined documentation. In addition, an API/library is under development to enable tools to “speak” CASE.

Acknowledgements
We are grateful to Sean Barnum for his work to promote standardization across cyber domains, and for generating the illustrative examples of data markings used in this paper (Section 5 - Formalizing obligations and controls for information sharing). In addition, we are grateful to the community that continues to develop CASE and UCO.

References

[1] W. Alink, R. Bhoedjang, P. Boncz, A. de Vries, (2006), *XIRAF - XML-based indexing and querying for digital forensics*, Proceedings of DFRWS2006, Journal of Digital Investigation, Volume 3 (Supplement): S50-8, Elsevier.

[2] ASTM (2015) *Standard Terminology for Digital and Multimedia Evidence Examination*, ASTM E2916 – 13.

[3] H.M.A van Beek, E.J. van Eijk, R.B. van Baar, M. Ugen, J.N.C. Bodde, A.J. Siemelink, (2015), *Digital forensics as a service: Game on*, Journal of Digital Investigation, Volume 15: 20–38, Elsevier

[4] R.A.F. Bhoedjang, A. R. van Ballegooijb, H.M.A van Beeka, J.C. van Schiea JC, F.W. Dillemab, R.B. van Baara, (2012), *Engineering an online computer forensic service*. Journal of Digital Investigation, Volume 9: 2, Elsevier.

[5] E. Casey, G. Back, S. Barnum, (2015), *Leveraging CybOX to standardize representation and exchange of digital forensic information,* Proceedings of DFRWS EU 2015, Journal of Digital Investigation, Volume 12 (Supplement), Elsevier.

[6] Casey E, Barnum S, Griffith R, Snyder J, van Beek H, Nelson A (in press), *The Evolution of Expressing and Exchanging Cyber-investigation Information in a Standardized Form*, In EU EVIDENCE Project publication, Springer.

[7] E. Casey, S. Barnum, R. Griffith, J. Snyder, H. van Beek, A. Nelson (in press), *Advancing Coordinated Cyber-investigations and Tool Interoperability using a Community Developed Specification Language*. Journal of Digital Investigation, Elsevier.

[8] M. Cohen, B. Schatz, S. Garfinkel, (2009), *Extending the advanced forensic format to accommodate multiple data sources, logical evidence, arbitrary information and forensic workflow*, Proceedings of DFRWS2009, Journal of Digital Investigation, Volume 6 (Supplement): S57-68. Elsevier.

[9] European Union Agency for Network and Information Security (2015), *Actionable information for security incident response* (https://www.enisa.europa.eu/publications/actionable-information-for-security)

[10] European Union Agency for Network and Information Security (2015), *Standards and tools for exchange and processing of actionable information* (https://www.enisa.europa.eu/publications/standards-and-tools-for-exchange-and-processing-of-actionable-information)

[11] S.L. Garfinkel, (2009), *Automating disk forensic processing with SleuthKit. In: XML and Python, systematic approaches to digital forensics engineering*, IEEE/SADFE 2009 Conference, Oakland, California; 2009.

[12] S.L. Garfinkel, (2012), *Digital forensics XML and the DFXML toolset*, Journal of Digital Investigation, Volume 8 (3-4): 161-74, Elsevier.

[13] S.L. Garfinkel, (2012), *Cross-drive analysis*, Proceedings of DFRWS2006, Journal of Digital Investigation 2012, Volume 3 (Supplement): S71-81. Elsevier.

[14] ISO/IEC 27037 (2012), *Information technology – Security techniques – Guidelines for identification, collection, acquisition and preservation of digital evidence*.

[15] ISO/IEC 27042 (2015), *Information technology – Security techniques – Guidelines for the analysis and interpretation of digital evidence*.

[16] B.N. Levine, M. Liberatore, (2009), *DEX: digital evidence provenance supporting reproducibility and comparison*, Journal of Digital Investigation, Volume 6: 48- 56, Elsevier. (Online https://github.com/umass-forensics/DEX-forensics)

[17] P. Turner, (2005), *Digital provenance - interpretation, verification and corroboration*, Journal of Digital Investigation, Volume 2: Pages 45-49 Elsevier.

[18] Office of the Director of National Intelligence, (2015), *XML Data Encoding Specification for Intelligence Document and Media Exploitation* (https://www.dni.gov/index.php/who-we-are/organizations/ic-cio/ic-cio-related-menus/ic-cio-related-links/ic-technical-specifications/document-and-media-exploitation).